Table of Contents

Storage Components

File Systems

Pleiades Home Filesystem

DRAFT

This article is being reviewed for completeness and technical accuracy.

The home file system on Pleiades (/u/username) is an SGI NEXIS 9000 filesystem. It is NFS-mounted on all of the Pleiades front-ends, bridge nodes and compute nodes.

Once a user is granted an account on Pleiades, the home directory is set up automatically during his/her first login.

Quota and Policy

Disk space quota limits are enforced on the home filesystem. By default, the soft limit is 8GB and the hard limit is 10GB. There are no inode limits on the home filesystem.

To check your quota and usage on your home filesystem, do:

```
%quota -v
Disk quotas for user username (uid xxxx):
     Filesystem  blocks   quota   limit   grace   files   quota   limit   grace
saturn-ib1-0:/mnt/home2
                7380152  8000000 40000000          190950       0       0
```

The quota policy for NAS states that if you exceed the soft quota, an email will be sent to inform you of your current usage and how much of your grace period remains. It is expected that a user will occasionally exceed their soft limit as needed, however after 14 days, users who are still over their soft limit will have their batch queue access to Pleiades disabled. If you believe that you have a long-term need for higher quota limits, you should send an email justification to support@nas.nasa.gov. This will be reviewed by the HECC Deputy Project Manager, Bill Thigpen, for approval.

The quota policy for NAS can be found here.

Backup Policy

Files on the home filesystem are backed up daily.

Pleiades Lustre Filesystems

Pleiades has several Lustre filesystems (/nobackupp[10-60]) that provide a total of about 3 PB of storage and serve thousands of cores. These filesystems are managed under Lustre software version 1.8.2.

Lustre filesystem configurations are summarized at the end of this article.

Which /nobackup should I use?

Once you are granted an account on Pleiades, you will be assigned to use one of the Lustre filesystems. You can find out which Lustre filesystem you have been assigned to by doing the following:

```
pfe1% ls -l /nobackup/your_username
lrwxrwxrwx 1 root root 19 Feb 23  2010 /nobackup/username -> /nobackupp30/username
```

In the above example, the user is assigned to /nobackupp30 and a symlink is created to point the user's default /nobackup to /nobackupp30.

TIP: Each Pleiades Lustre filesystem is shared among many users. To get good I/O performance for your applications and avoid impeding I/O operations of other users, read the articles: Lustre Basics and Lustre Best Practices.

Default Quota and Policy on /nobackup

Disk space and inodes quotas are enforced on the /nobackup filesystems. The default soft and hard limits for inodes are 75,000 and 100,000, respectively. Those for the disk space are 200GB and 400GB, respectively. To check your disk space and inodes usage and quota on your /nobackup, use the *lfs* command and type the following:

```
%lfs quota -u username /nobackup/username
Disk quotas for user username (uid xxxx):
   Filesystem  kbytes      quota   limit   grace   files   quota   limit   grace
/nobackup/username 1234  210000000 420000000    -     567   75000  100000       -
```

The NAS quota policy states that if you exceed the soft quota, an email will be sent to inform you of your current usage and how much of your grace period remains. It is expected that users will occasionally exceed their soft limit, as needed; however after 14 days, users who are still over their soft limit will have their batch queue access to Pleiades disabled.

If you anticipate having a long-term need for higher quota limits, please send a justification via email to support@nas.nasa.gov. This will be reviewed by the HECC Deputy Project Manager for approval.

For more information, see also, Quota Policy on Disk Space and Files.

NOTE: If you reach the hard limit while your job is running, the job will die prematurely without providing useful messages in the PBS output/error files. A Lustre error with code -122 in the system log file indicates that you are over your quota.

In addition, when a Lustre filesystem is full, jobs writing to it will hang. A Lustre error with code -28 in the system log file indicates that the filesystem is full. The NAS Control Room staff normally will send out emails to the top users of a filesystem asking them to clean up their files.

Important: Backup Policy

As the names suggest, these filesystems are not backed up, so any files that are removed *cannot* be restored. Essential data should be stored on Lou1-3 or onto other more permanent storage.

Configurations

In the table below, /nobackupp[10-60] have been abbreviated as p[10-60].

Pleiades Lustre Configurations

Filesystem	p10	p20	p30	p40	p50	p60
# of MDSes	1	1	1	1	1	1
# of MDTs	1	1	1	1	1	1
size of MDTs	1.1T	1.0T	1.2T	0.6T	0.6T	0.6T
# of usable inodes on MDTs	~235x10^6	~115x10^6	~110x10^6	~57x10^6	~113x10^6	~123x10^6
# of OSSes	8	8	8	8	8	8
# of OSTs	120	60	120	60	60	60
size/OST	7.2T	7.2T	3.5T	3.5T	7.2T	7.2T
Total Space	862T	431T	422T	213T	431T	431T
Default Stripe Size	4M	4M	4M	4M	4M	4M
Default Stripe Count	1	1	1	1	1	1

NOTE: The default stripe count and stripe size were changed on January 13, 2011. For directories created prior to this change, if you did not explictly set the stripe count and/or stripe size, the default values (stripe count 4 and stripe size 1MB) were used. This means that files created prior to January 13, 2011 had those old default values. After this date, directories without an explicit setting of stripe count and/or stripe size adopted the new stripe count of 1 and stripe size of 4MB. However, the old files in that directory will retain their old default values. New files that you create in these directories will adopt the new default values.

Columbia Home Filesystems

DRAFT

This article is being reviewed for completeness and technical accuracy.

Columbia's home fileystem (/u/username) is NFS-mounted on the Columbia front-end (cfe2) and compute nodes (Columbia21-24).

Once a user is granted an account on Columbia, the home directory is set up automatically during his/her first login.

Quota and Policy

Disk space quota limits are enforced on the home filesystem. By default, the soft limit is 4GB and the hard limit is 5GB. There are no inode limits on the home filesystem.

To check your quota and usage on your home filesystem, do:

```
%quota -v
Disk quotas for user username (uid xxxx):
     Filesystem  blocks   quota   limit   grace   files   quota   limit   grace
  ch-rg1:/home6    4888  4000000 5000000            294       0       0
```

The quota policy for NAS states that if you exceed the soft quota, an email will be sent to inform you of your current usage and how much of your grace period remains. It is expected that a user will occasionally exceed their soft limit as needed; however after 14 days, users who are still over their soft limit will have their batch queue access to Pleiades disabled. If you believe that you have a long-term need for higher quota limits, you should send an email justification to support@nas.nasa.gov. This will be reviewed by the HECC Deputy Project Manager, Bill Thigpen, for approval.

The quota policy for NAS can be found here.

Backup Policy

Files on the home filesystem are backed up daily.

Columbia CXFS Filesystems

Columbia CXFS filesystems (/nobackup[1-2][a-i]) are shared and accessible from cfe2 and Columbia21-24. This allows user jobs to be load-balanced across Columbia's systems without forcing users to move their data to a particular Columbia system.

Users will have a nobackup directory on one of these shared file systems. To find out where your nobackup directory is, log in to the front-end node and type the following shell command:

```
cfe2% ls -d /nobackup[1-2][a-i]/$USER
/nobackup1f/username/
```

In this example, the user is assigned to /nobackup1f.

Default Quota and Policy on /nobackup

Disk space and inodes quotas are enforced on the CXFS /nobackup[1-2][a-i] filesystems. The default soft and hard limits for inodes are 25,000 and 50,000, respectively. Those for disk space are 200GB and 400GB, respectively. To check your disk space and inodes usage and quotas on your CXFS filesystem, do the following:

```
cfe2% quota -v
Disk quotas for user username (uid xxxx):
     Filesystem  blocks   quota   limit   grace   files   quota   limit   grace
/dev/cxvm/nobackup1f
                1673856  210000000 420000000              10973   25000   50000
```

The NAS quota policy states that if you exceed the soft quota, an email will be sent to inform you of your current usage and how much of your grace period remains. It is expected that users will occasionally exceed their soft limit, as needed; however after 14 days, users who are still over their soft limit will have their batch queue access to Columbia disabled.

If you anticipate having a long-term need for higher quota limits, please send a justification via email to support@nas.nasa.gov. This will be reviewed by the HECC Deputy Project Manager for approval.

For more information, see also, Quota Policy on Disk Space and Files.

Important: Backup Policy

As the names suggest, these filesystems are not backed up, so any files that are removed *cannot* be restored. Essential data should be stored on Lou1-3 or onto other more permanent storage.

Accessing CXFS from Lou

The Columbia CXFS filesystems are also mounted on Lou1-3. This allows you to copy files between the CXFS filesystems and your Lou home filesystem, using the *cp* or *cxfscp* commands on Lou.

Archive Systems

Mass Storage Systems: Lou1 and Lou2

The NAS environment contains three mass storage systems, Lou1 and Lou2, to provide long-term data storage for users of our high-end computing systems. These storage systems are SGI Altix computers running the Linux operating system. The disk space for the three systems combined is about 290 terabytes (TB), which is split into filesystems ranging from 9-30 TB in size.

Which Lou System I Should Use?

Each user should be able to log into any of the Lou systems, but will only have storage space on the home filesystem of one of them. Follow the steps below to determine which system you should store data on.

1. Log in to either Lou1 or Lou2. For example:

   ```
   your_localhost% ssh nas_username@lou1.nas.nasa.gov
   ```

2. Type the command "mylou" to find out your mass storage host. For example:

   ```
   lou1% mylou

   Your Mass Storage host is lou2

   Store files there in your home directory, /u/your_nas_username
   ```

   Be aware that Lou1 and Lou2 do *not* share their home filesystems.

3. Use the home filesystem on Lou*X* (where *X* = 1 or 2) determined by the step above for your long-term storage. For example:

   ```
   pfe1% scp foo lou2:
   ```

Quota Limits On Lou

For Lou*X* (where *X* = 1 or 2) that is assigned to you, there are no disk quota limits on your home filesystem. On the other hand, there *are* limits on the number of files (inode):

- 250,000 inode soft limit (14-day grace period)
- 300,000 inode hard limit

See Policy on Disk Files Quotas for Lou for more information.

Data (Un)Migration Between Disk and Tapes

In addition to the disk space, Lou1 and 2 have a combined 64 LTO-4 tape drives. Each of the LTO-4 tapes holds 800 GB of uncompressed data. The total storage capacity is approximately 10 PB.

Data stored on Lou's home filesystems (disk) is automatically migrated to tapes whenever necessary to make space for more data. Two copies of your data are written to tape media in silos located in separate buildings.

Data migration (from disk to tape) and unmigration (from tape to disk) are managed by the SGI Data Migration Facility (DMF) and Tape Management Facility (TMF).

If you need some data that is only available on tapes, make sure to unmigrate the data from tape to your home filesystem on Lou before transferring it to other systems.

For more tips on how to use Lou more effectively, see Storage Best Practices.

Network

TCP Performance Tuning for WAN Transfers

DRAFT

This article is being reviewed for completeness and technical accuracy.

The purpose of this document is to help you maximize your wide-area network bulk data transfer performance by tuning the TCP settings for your end hosts. These are some common configuration tasks for enabling high performance data transfers on your system.

Making changes to your system should only be done by a lead system administrator or someone who is authorized to make changes.

Linux

1. Edit */etc/sysctl.conf* and add the following lines:

   ```
   net.core.wmem_max = 4194304
   net.core.rmem_max = 4194304
   ```

2. Then have them loaded by running "sysctl -p".

Windows

We recommend using a tool like Dr. TCP

Set the "Tcp Receive Window" to at least 4000000, turn on "Window Scaling", "Selective Acks", and "Time Stamping".

Other options for tuning Windows XP TCP are the SG TCP Optimizer or using Windows Registry Editor to edit the registry, but this is only recommended for Windows users who are already familiar with registry parameters.

Mac

- **Do these steps for OS 10.4**

These changes require root access.

In order to allow the Mac operating system to retain the parameters after a reboot, edit the following variables in */etc/sysctl.conf*:

```
# Set maximum TCP window sizes to 4 megabytes

net.inet.tcp.sendspace= 4194304
net.inet.tcp.recvspace= 4194304
# Set maximum Socket Buffer sizes to 4 megabytes

kern.ipc.maxsockbuf= 4194304
```

- **Do these steps for OS 10.5 and up**

  Use the **sysctl** command for the following variable:

  ```
  sysctl -w net.inet.tcp.win_scale_factor=8
  ```

If you follow these steps and are still getting less than your expected throughput, please contact the network group at support@nas.nasa.gov attn: Networks and we will work with you on tuning your system to optimize file transfers. You can also try the additional steps outlined in the following documents: Optional Advanced Tuning For Linux and Tips for File Transfers.

Optional Advanced Tuning for Linux

DRAFT

This article is being reviewed for completeness and technical accuracy.

This document describes additional TCP settings that can be tuned on high performance Linux systems. This is intended for 10 Gigabit hosts, but can also be applied to 1 Gigabit hosts. The following steps should be taken in addition to the steps outlined in TCP Performance Tuning for WAN transfers.

Configure the following */etc/sysctl.conf* settings for faster TCP:

Set maximum TCP window sizes to 12 megabytes

```
net.core.rmem_max = 11960320
net.core.wmem_max = 11960320
```

Set minimum, default, and maximum TCP buffer limits

```
net.ipv4.tcp_rmem = 4096 524288 11960320
net.ipv4.tcp_wmem = 4096 524288 11960320
```

Set maximum network input buffer queue length

```
net.core.netdev_max_backlog = 30000
```

Disable caching of TCP congestion state (Linux Kernel version 2.6 only). Fixes a bug in some Linux stacks.

```
net.ipv4.tcp_no_metrics_save = 1
```

Use the BIC TCP congestion control algorithm instead of the TCP Reno algorithm, Linux Kernel versions 2.6.8 to 2.6.18

```
net.ipv4.tcp_congestion_control = bic
```

Use the CUBIC TCP congestion control algorithm instead of the TCP Reno algorithm, Linux Kernel versions 2.6.18+

```
net.ipv4.tcp_congestion_control = cubic
```

Set the following to 1 (should default to 1 on most systems):

```
net.ipv4.tcp_window_scaling =1
net.ipv4.tcp_timestamps = 1
net.ipv4.tcp_sack = 1
```

A reboot will be needed for changes to */etc/sysctl.conf* to take effect, or you can attempt to reload sysctl settings (as root) with 'sysctl -p'.

For additional information visit this web site

If you have a 10Gig system or if you follow these steps and are still getting less than your expected throughput, please contact support@nas.nasa.gov attn: Networks and we will work with you on tuning your system to optimize file transfers.

NAS VPN Service

DRAFT

This article is being reviewed for completeness and technical accuracy.

For remote users wishing to connect to limited resources available on the local NAS network, a virtual private network service is available to any existing NAS user who has a Lou account and active SecurID fob. Additionally, NAS support staff may also make use of this service to provide some remote support to your government systems while on travel or at home.

*** ALL system traffic is routed through NAS while you are connected via VPN ***

This system is intended for government use and users are required to follow the appropriate use policy. All traffic is monitored. While connected, **ALL** your traffic will be routed through NAS, as if you were physically connected to NASLAN. When you are finished with your session, please remember to log out.

Users are subject to the NAS VPN Security Policy.

When and Why to use VPN:

The VPN service will make your system appear to be logically within the NAS external network, with some access to internal resources. Connection to the VPN server and installation of the VPN software are handled through Javascript and users do not have to pre-install prior to connection. VPN makes use of your standard web browser for encryption.

Through VPN, users can make use of any of the following services:

- Apple file sharing (AFP remote mount)
- Access to the NAS license server
- Access to the internal NAS webserver
- Access to some ARC websites not accessible from the outside
- SSH and SCP directly into NASLAN systems (NOT to Enclave systems)

How to Login to NAS VPN:

The login site is at: https://nas-vpn.nas.nasa.gov

Enter your login credentials (remember, it is your Lou account information and SecurID fob).

Click "Sign In", and you will be taken to the default VPN page upon successful authentication.

If it is the first time connecting from a new system, you will be prompted to install some software. Depending on your browser's security level, you may get a banner on top of the page warning you that the site is trying to install an ActiveX control. Click on that banner to install it.

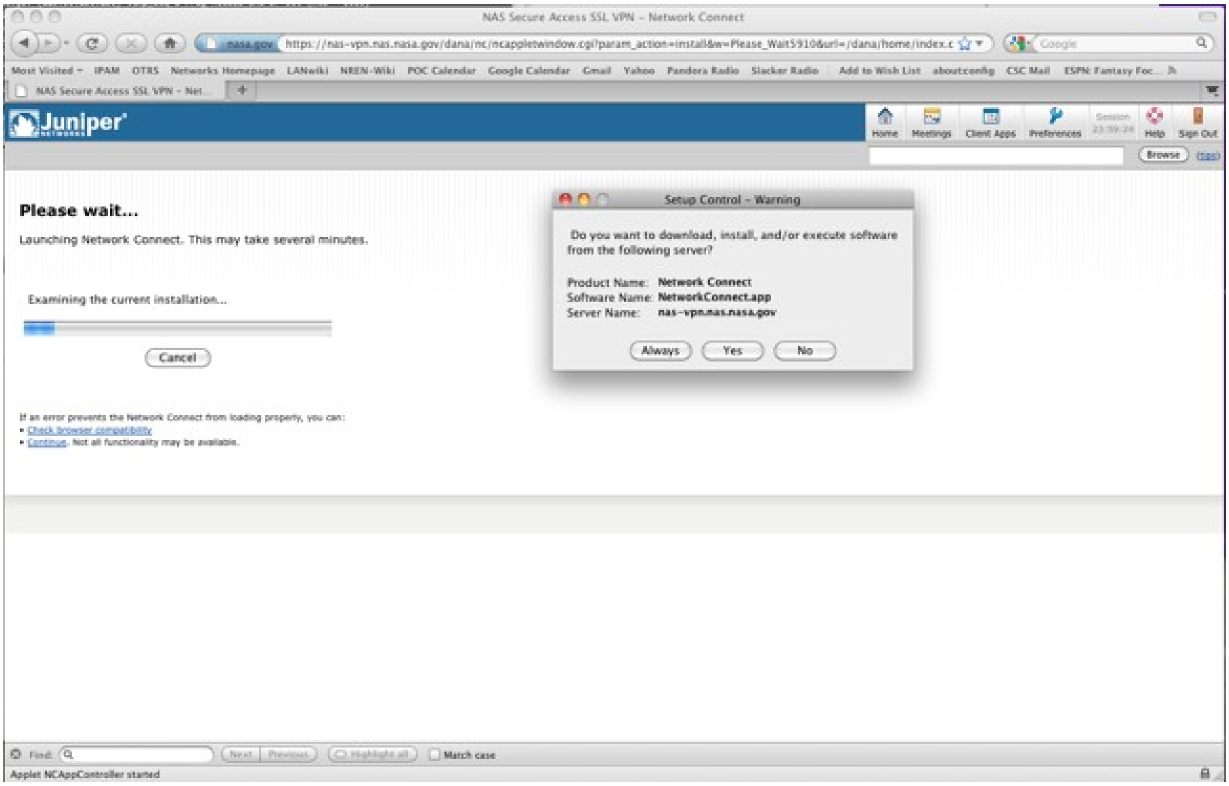
A pop-up window will show that the connection is negotiating. If it does not automatically start, click on "Start" under the Client Application Session for Network Connect. This may take up to a minute. Once done, the window will close and you will have been assigned a new IP address inside the VPN environment. A little icon will be placed in your desktop tray which you can use to get session information and disconnect.

You can verify that the VPN is working by connecting to:

http://myipaddress.com

Your address should be on the 198.9.30.x network.

NAS Supported systems already have the VPN Client installed on them. Just double click the "Network Connect" icon and enter your authentication credentials.



NAS Secure Access SSL VPN – Network Connect

https://nas-vpn.nas.nasa.gov/dana/nc/ncappletwindow.cgi?param_action=install&w=Please_Wait5910&url=/dana/home/index.c

Please wait...

Launching Network Connect. This may take several minutes.

Examining the current installation...

If an error prevents the Network Connect from loading properly, you can:
- Check browser compatibility
- Continue. Not all functionality may be available.

Setup Control – Warning

Do you want to download, install, and/or execute software from the following server?

Product Name: Network Connect
Software Name: NetworkConnect.app
Server Name: nas-vpn.nas.nasa.gov

Applet NCAppController started

Alternatively, if you cannot connect through the above methods, you can manually download the VPN client software here:

- Download VPN client for Mac
- Download VPN client for Linux
- Download the tar file which contains the VPN client for Windows
- Download the tar file which contains the VPN client for Windows 64 bit systems

VPN FAQ's:

- DO I need to keep my web browser open to keep the VPN up?

  No, you do not need to keep your web browser open once you start up the Network Connect client.
- How do I disconnect from the VPN?

  You can either go back to http://nas-vpn.nas.nasa.gov and click the "Sign Out" button on the top right corner, or right-click on the icon in your system tray and select "Sign Out".
- Is there an auto-logout?

  Yes. You will be logged out after 30 minutes of inactivity. The max session length is 12 hrs before you need to re-authenticate and you will be given a reminder before being disconnected. This is so that people don't stay "camped" on the VPN network.
- What traffic is sent over the VPN?

  ALL traffic you send will be sent over the VPN. This includes any websites you visit, any chat programs, or any software that requires a network connection. Because of this, it is important that you disconnect from the VPN while not in use.
- What changes are made to my system?

  Several changes are made to your network including a new IP address, default route, search domain and other minor files which allow you to be "virtually" inside the NASLAN. This means you can refer to hosts just by their hostname and not their fully-qualified name - eg:

  ssh username@desktop
- What browsers are supported?

  As long as you meet the requirements listed above, you should be able to connect on Safari, Internet Explorer, and Firefox. We recommend you update to the latest stable version. The minimum browser requirements are:

  - ♦ 168-bit and greater encryption
  - ♦ SSLv3 and TLSv1
  - ♦ JRE / Java enabled
  - ♦ Pop-up Windows

- How will connecting to the VPN impact my home network?

Some of your home services may stop working while connected to the VPN. This includes services like Internet printing, file sharing, and audio streaming. This is to ensure security of NAS while you are connected to the VPN.

NAS Remote Network Diagnostic Tools

A NAS network service that enables remote users to test end-to-end connectivity to the NAS HECC enclave is now available. Users can access the Network Diagnostic Tool service and initiate tests at: http://npad.nas.nasa.gov.

The diagnostic tests can only run on a Java-enabled web browser. If you have trouble accessing the website, please contact the NAS Control room (see below) and we will assist you.

Features

- Tools are accessible from any standard web browser
- Command-line tools are also available for Linux servers
- Users receive a diagnostic report on the test results; a copy of the report is sent to the NAS Networks team for analysis. If any problems are identified, the team will contact you to help resolve the issue.

The services available on this website run from inside the NAS network and are connected at 10 Gigabit Ethernet rates.

The Network Diagnostic Tester (NDT) - Performs a quick test that reports the maximum throughput to your remote system from NAS. It will also identify any issues with possible bad cabling, negotiation issues, packet loss, or general network congestion.

Network Path Application Diagnosis (NPAD) tool - Performs a more elaborate connection test and determines problems with TCP parameters, buffer sizes, and/or router congestion, and notifies you of recommended settings for maximum performance.

Traceroute tool - Allows users to perform reverse traceroutes from NAS display the path and measure transit delays of packets to a remote host.

Ping tool - Allows users to perform reverse pings from NAS to the remote host, to test the reachability to your host and measure round-trip time for messages to reach the remote host.

If you want a command line interface, you can also download client software and link to various other public services from this website.

If you have any questions about these new services, please contact the NAS Control Room staff 24x7 at (800) 331-8737, (650) 604-4444, support@nas.nasa.gov.

Increasing File Transfer Rates

One challenge users face is moving large amounts of data efficiently to/from NAS across the network. Often, minor system, software, or network configuration changes can increase network performance an order of magnitude or more. This article describes some methods for increasing data transfer performance.

If you are experiencing slow transfer rates, try these quick tips:

- Transfer using the bridge nodes (bridge1, bridge2) instead of the Pleiades front-end systems (PFEs). The bridge nodes have much more memory, along with 10-Gigabit Ethernet interfaces to accommodate many large transfers. The PFEs often become oversubscribed and cause slowness.
- If using the scp command, make sure you are using OpenSSH version 5 or later. Older versions of SSH have a hard limit on transfer rates and are not designed for WAN transfers. You can check your version of SSH by running the command ssh -V.
- For large files that are a gigabyte or larger, we recommend using BBFTP. This application allows for transferring simultaneous streams of data and doesn't have the overhead of encrypting all the data (authentication is still encrypted).

Online Network Testing Tools

The NAS PerfSONAR Service provides a custom website that that allows you to quickly self-diagnose your remote network connection issues, and reports the maximum bandwidth between sites, as well as any problems in the network path. Command-line tools are available if your system does not have a web browser.

Test results are also sent to our network experts, who will analyze traffic flows, identify problems, and work to resolve any bottlenecks that limit your network performance, whether the problem is at NAS or a remote site.

One-on-One Help

If you still require assistance in increasing your file transfer rates, please contact the NAS Control Room at support@nas.nasa.gov, and a network expert will work with you or your local administrator one-on-one to identify methods for increasing your rates.

To learn about other network-related support areas. see also, End-to-End Networking Services.

Auxiliary Systems

Bridge nodes

DRAFT

This article is being reviewed for completeness and technical accuracy.

Currently, the Pleiades Lustre filesystems are not mounted on Lou. File transfers between Pleiades and Lou are normally done with remote file transfer commands such as scp, bbftp and bbscp.

Using the Pleiades bridge nodes, one can actually transfer files between Pleiades' home or Lustre filesystems and Lou's home filesystem through Columbia's CXFS filesystems.

In the example below, on bridge1, the file *foo* is copied from a user's Pleiades Lustre /nobackup filesystem to his Columbia CXFS filesystem under /nobackup2a. Then, on Lou, the same file is copied from /nobackup2a to user's Lou home filesystem.

```
bridge1% cp /nobackup/username/foo /nobackup2a/username
```

```
lou1% cp /nobackup2a/username/foo /u/username
```